

Metadata Portal Project: Using DDI to Enhance Data Access and Dissemination

Mary Vardigan
Assistant Director, ICPSR
Director, DDI Alliance



Metadata Portal Project

- NSF-funded project to improve metadata for the General Social Survey and American National Election Study series (SES-1229957)
- Several partners



More on the Project

- Showcases what you can do with rich structured metadata
- Products designed to apply to other survey data
- Intended to encourage more software development
- Spans two years with three phases
 - Create markup
 - Establish portal with prototype tools
 - Revise business process to capture metadata and reduce cost

Metadata Portal Project -- Phase 1

Deliverables:

- Inventory of all ANES, ICPSR, and GSS files
 - Resolve any versioning issues
- Markup (DDI Codebook level) of all existing ANES and GSS codebooks, shared freely
- Profiles of markup for specific purposes
 - Indicate which elements are needed

Metadata Portal Project -- Phase 1 (cont.)

- Markup enhancements (DDI Lifecycle) for a subset of the codebooks, to include:
 - Provenance and process (e.g., recodes)
 - Harmonization and comparison (e.g., concepts)
- “Reverse universe generator” to detect and mark up branching patterns and missing data
- Bibliographic references
- Findings and lessons learned

Metadata Portal Project -- Phase 2

Deliverables:

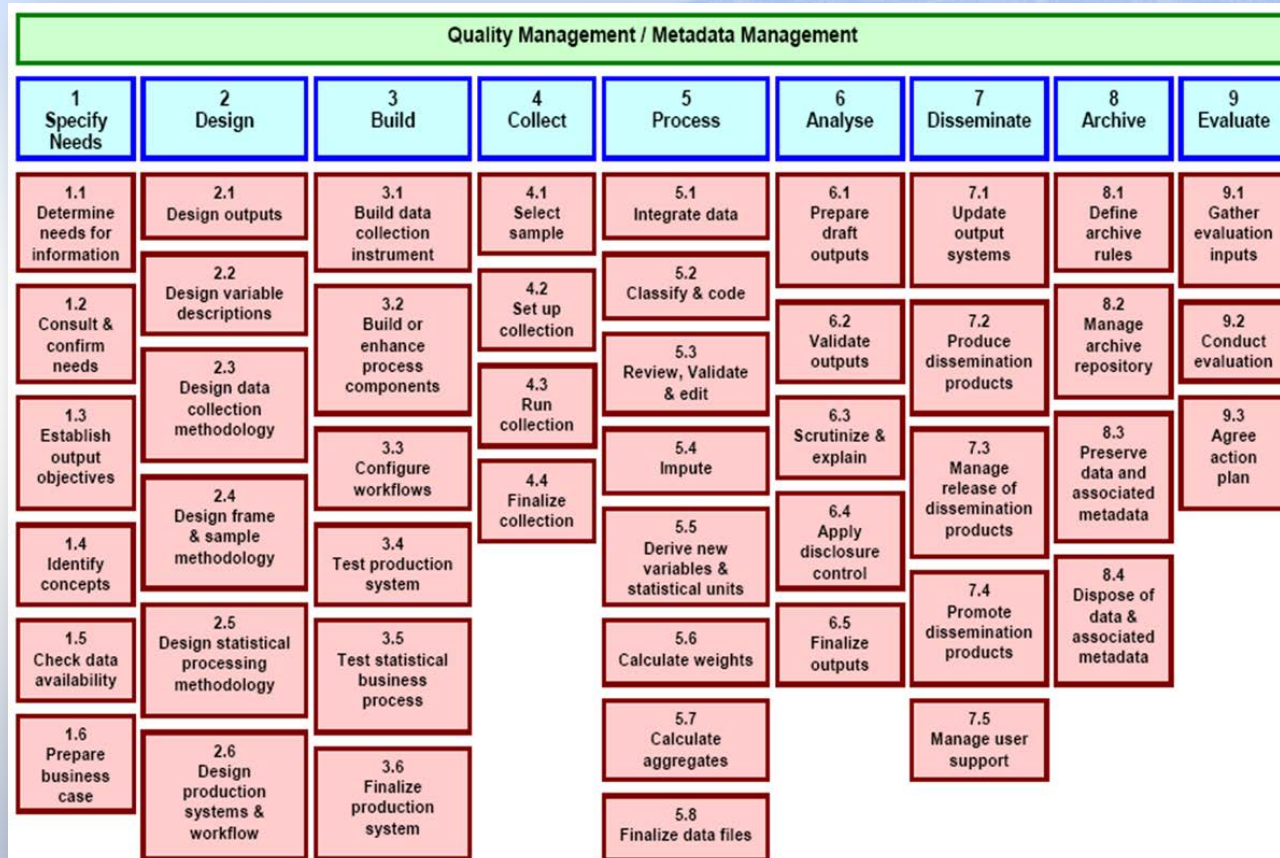
- A Metadata Portal demonstrating tools such as:
 - Survey catalog and documentation library
 - Question banks
 - Classification banks
 - Dynamic codebook generators
 - Script generators for syntax
 - Variable comparison tools
 - Resources for developers

Metadata Portal Project -- Phase 3

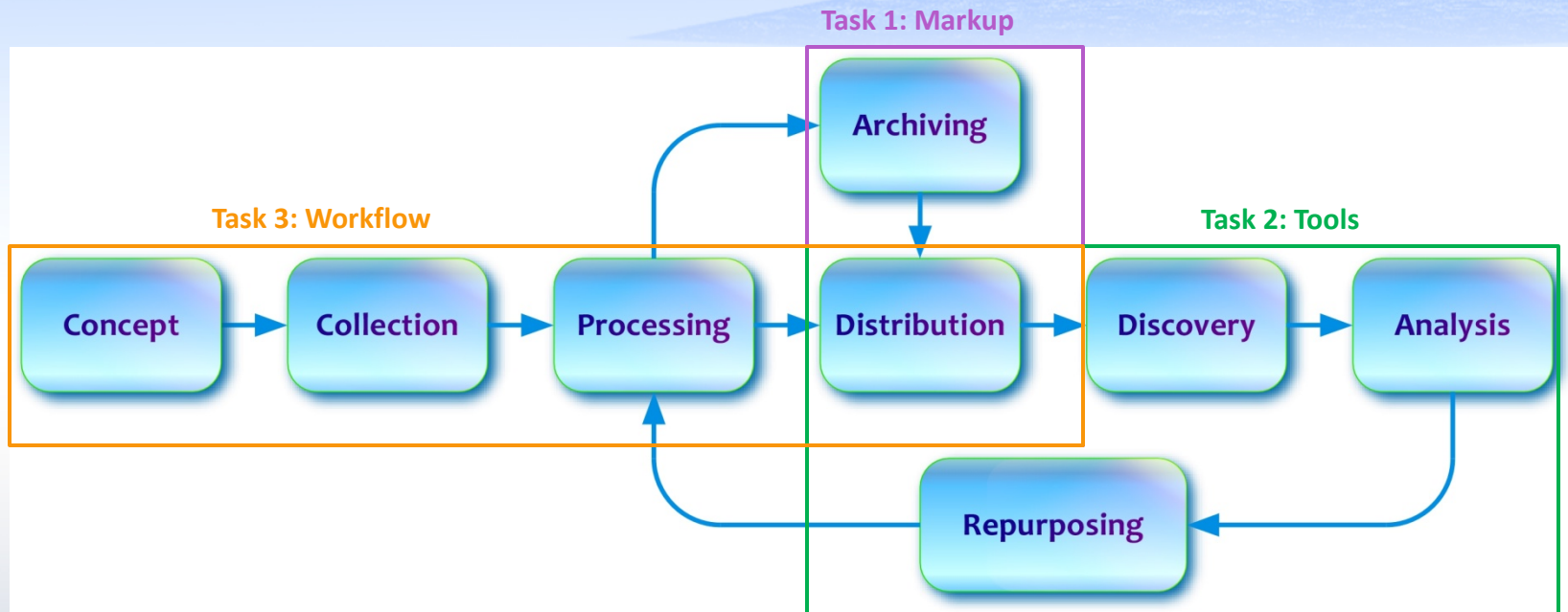
Deliverables:

- For both GSS and ANES, develop new metadata-driven workflows to capture metadata early in the data collection process for re-use
- Disseminate findings to others
- GSS has already started mapping its processes to the Generic Statistical Business Process Model (GSBPM)

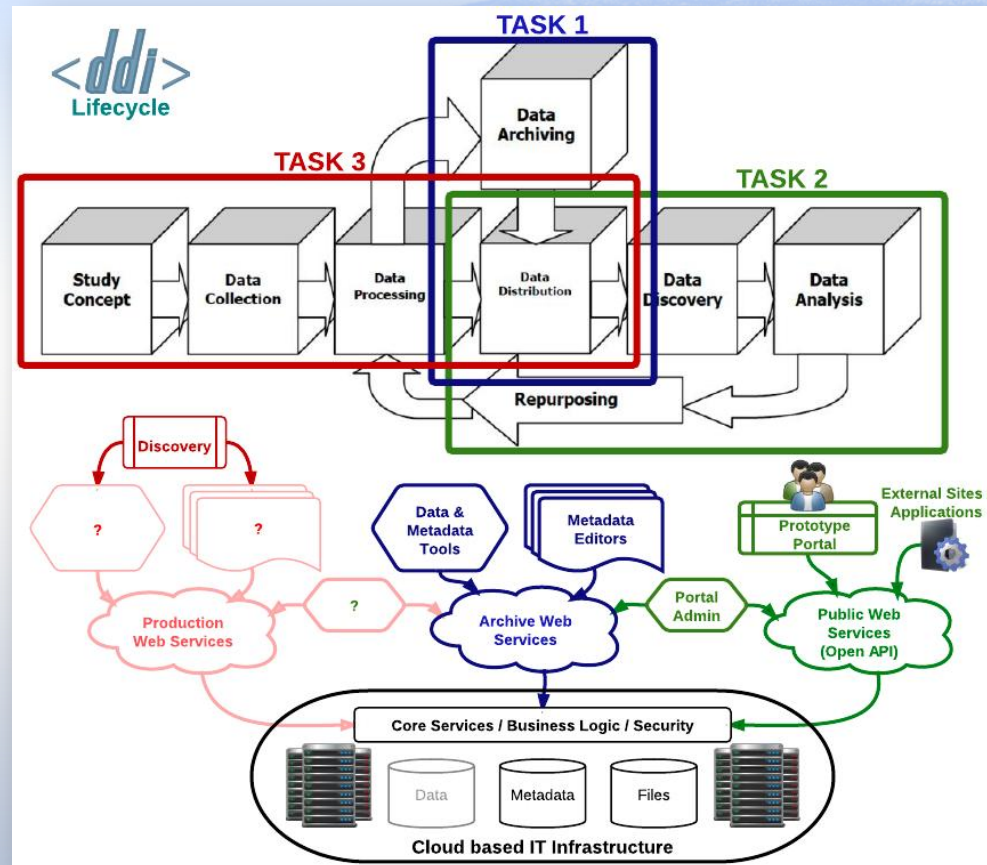
GSBPM



How Project Relates to Data Lifecycle



Project Infrastructure



Where Are We Now?

- Inventorying and analyzing all ANES and GSS files ported to iRODS repository
- Populating a spreadsheet with study-level information from all three data sources
- Getting started on the reverse-engineering project to understand routing through questionnaire
- Generating variable-level DDI markup

Where Are We Going?

- Looking to create innovative tools based on structured metadata
- Some ICPSR examples...

Search for Variables

Variable Search Results

Your query returned 20 variables. Use the search box below to revise your search , or [start a new search](#) .

afghanistan war worth it

REVISE SEARCH

Results 1 - 20

Results per page: 25

Sort by: Relevance

[1]

Compare

Name Label/Question

Clear All



[V043131](#) [M1a. War in Afghanistan worth the cost](#)

Taken from: American National Election Study, 2004: Pre- and Post-Election Survey.



[F1ZF2](#) [f1zf2. Will Afghanistan war have been worth the costs](#)

When all is said and done, do you think the U.S. involvement in the Afghanistan war will have been worth the costs, or will it not have been worth the costs? ... ([view details](#))

Taken from: American National Election Studies (ANES) Panel Recontact Study, 2010.



[V023120](#) [Q4a4. War in Afghanistan Worth Cost](#)

Q4a4. Taking everything into account, do you think the U.S. war against the Taliban government in Afghanistan was WORTH THE COST or NOT? ... ([view details](#))

Taken from: American National Election Study, 2002: Pre- and Post-Election Survey - American National Election Study, 2002: Pre- and Post-Election Survey.



[P023120](#) [Q4a4. War in Afghanistan Worth Cost](#)

Q4a. IF PANEL RESPONDENT: After the September 11 terrorist attacks, President Bush declared a war on terrorism. A first step was to launch air strikes against the Taliban government of Afghanistan that was providing aid and protection to O... ([view details](#))

Check Variables to Compare



C1_ZL2 [c1_zl2. When all is said and done, do you think the U.S. involvement in the Afghanistan war will have been worth the costs, or will it not have been worth the costs?](#)

When all is said and done, do you think the U.S. involvement in the Afghanistan war will have been worth the costs, or will it not have been worth the costs? ... ([view details](#))

Taken from: American National Election Studies: Evaluations of Government and Society Study 1 (EGSS 1), 2010-2012.



Q14 [Q.14 Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not?](#)

Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not? ... ([view details](#))

Taken from: ABC News/Washington Post Poll, February 2007.



Q15 [Q.15 Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not?](#)

Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not? ... ([view details](#))

Taken from: ABC News/Washington Post Monthly Poll, July 2008.

Display Side by Side

Compare Variables

[< Back to search results](#)

☒ C1_ZL2

[American National Election Studies: Evaluations of Government and Society Study 1 \(EGSS 1\), 2010-2012](#)

c1_zl2. When all is said and done, do you think the U.S. involvement in the Afghanistan war will have been worth the costs, or will it not have been worth the costs?

Question

When all is said and done, do you think the U.S. involvement in the Afghanistan war will have been worth the costs, or will it not have been worth the costs?

Responses

| Value | Label | Unweighted Frequencies | % |
|-------|---------------------------------------|------------------------|------|
| 1 | 1. Will have been worth the costs | 332 | 26.8 |
| 2 | 2. Will not have been worth the costs | 835 | 67.3 |
| | Missing Data | | |
| -7 | -7. No answer | 29 | 2.3 |
| -6 | -6. Not asked, unit non-response | 10 | 0.8 |
| -5 | -5. Not asked, terminated | 34 | 2.7 |
| -2 | -2. Missing, see documentation | 0 | 0.0 |
| -1 | -1. Inapplicable | 0 | 0.0 |
| | Total | 1,240 | 100% |

Time Period

2010-10

Universe

United States citizens aged 18 years or older.

☒ Q14

[ABC News/Washington Post Poll, February 2007](#)

Q.14 Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not?

Question

Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not?

Responses

| Value | Label | Unweighted Frequencies | % |
|-------|------------------------|------------------------|------|
| 1 | Yes, worth fighting | 633 | 58.5 |
| 2 | No, not worth fighting | 417 | 38.5 |
| | Missing Data | | |
| -9 | Other (specify) | 0 | 0.0 |
| -7 | Don't know | 0 | 0.0 |
| -6 | Refused | 0 | 0.0 |
| -5 | No Answer | 0 | 0.0 |
| -4 | Blank | 0 | 0.0 |
| -1 | System Missing | 2 | 0.2 |
| 8 | DK/No opinion | 30 | 2.8 |
| | Total | 1,082 | 100% |

Time Period

2007-02

Universe

Persons aged 18 and over living in households with telephones in the contiguous 48 United States.

☒ Q15

[ABC News/Washington Post Monthly Poll, July 2008](#)

Q.15 Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not?

Question

Thinking now about Afghanistan, all in all, considering the costs to the United States versus the benefits to the United States, do you think the war in Afghanistan was worth fighting, or not?

Responses

| Value | Label | Unweighted Frequencies | % |
|-------|------------------------|------------------------|------|
| 1 | Yes, worth fighting | 584 | 52.2 |
| 2 | No, not worth fighting | 480 | 42.9 |
| | Missing Data | | |
| -9 | Other (specify) | 0 | 0.0 |
| -7 | Don't know | 0 | 0.0 |
| -6 | Refused | 0 | 0.0 |
| -5 | No Answer | 0 | 0.0 |
| -4 | Blank | 0 | 0.0 |
| -1 | System Missing | 4 | 0.4 |
| 8 | DK/No opinion | 51 | 4.6 |
| | Total | 1,119 | 100% |

Time Period

2008-07

Universe

Persons aged 18 and over living in households with telephones in the contiguous 48 United States.

Search for Studies by Variable Relevance

Study Search Results

Your query returned 4132 results. Use the search box below to revise your search, or [start a new search](#).

Results 1 - 25
Results per page: 25
Sort by: Variable Relevance (beta)
[1] 2 3 4 ▶ ▶

Study Title/Investigator

- [Drug Abuse Warning Network \(DAWN\), 1994: \[United States\]](#)** (ICPSR 2756)
United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Office of Applied Studies

53 variables match age including:
 - ☐ [AGE: AGE OF PATIENT](#) Specifies the age group of the patient at the time of the emergency department visit.
 - ☐ [DRUG01: ACETAMINOPHEN](#) Indicates each drug or drug category mentioned. These consist of the 50 drug categories most frequ
 - ☐ [DRUG02: ALCOHOL-IN-COMBO](#) Indicates each drug or drug category

52 variables match gender including:
 - ☐ [DRUG01: ACETAMINOPHEN](#) Indicates each drug or drug category mentioned. These consist of the 50 drug categories most frequ
 - ☐ [DRUG02: ALCOHOL-IN-COMBO](#) Indicates each drug or drug category mentioned. These consist of the 50 drug categories most frequ
 - ☐ [DRUG03: ALPRAZOLAM](#) Indicates each

157 variables match drug use including:
 - ☐ [MOTIVE: DRUG USE MOTIVE](#) Specifies the patient's motive for taking the drug, e.g., dependence, recreational use, suicide, et
 - ☐ [DRUG52: DRUGS, UNKNOWN](#) Indicates each drug or drug category mentioned. These consist of the 50 drug categories most frequ
 - ☐ [DRUG51: DRUGS, OTH CAT](#)

Variables in Context

ICPSR Find & Analyze Data

Find Data Search/Compare Variables Find Publications Resources for Students Get Help

NOTE: The variable list utility below is still in testing. Please report any problems encountered to web-support@icpsr.umich.edu.

Variable Groups

- [-] SURVEY/DATASET INFORMATION
- [-] DEMOGRAPHICS/RESPONDENT CHARACTERISTICS
 - [-] Geographic
 - [-] Respondent Characteristics
 - V148 042(R) :AGE <>18 DICHOTOMY**
 - V150 042C03 :R'S SEX
 - V151 042C04(R)R'S RACE
 - V152 042C05 :R SPD >TIM R-URB
 - V153 042C06 :R NOT MARRIED
 - [-] Household/Family
- [-] ATTITUDES AND BELIEFS
- [-] ASPIRATIONS
- [-] EMPLOYMENT/INCOME
- [-] RECREATION
- [-] DRIVING/SUBSTANCE ABUSE
- [-] SUBSTANCE USE

series: [Monitoring the Future \(MTF\) Series](#) > study: [Monitoring the Future: A Continuing Study of American Youth \(12th-Grade Survey\), 2004](#) > variable: V148 >

V148: 042(R) :AGE <>18 DICHOTOMY

Item Number: Component questions: 1) "In what year were you born?" (item 00010), 2) "In what month were you born?" (item 00020), and 3) "How many years old" 1="18 years of age and over"

| Value | Label | Unweighted Frequency | % |
|-------|--------------|----------------------|-------|
| 1 | < 18 | 6,648 | 43.7% |
| 2 | 18+ | 7,996 | 52.5% |
| | Missing Data | | |
| -9 | MISSING | 578 | 3.8% |
| | Total | 15,222 | 100% |

Based upon 14,644 valid cases out of 15,222 total cases.

Location: 125-126 (width: 2; decimal: 0)

Variable Type: numeric

(Range of) Valid Data Values: 1 thru 2

(Range of) Missing Values: -9

Notes

Source: This variable was taken from: [Monitoring the Future: A Continuing Study of American Youth \(12th-Grade Survey\), 2004](#).

All made
possible by
rich, structured
DDI metadata

```
<titl>American National Election Study, 2004: Pre- and Post-
Election Survey</titl>
<IDNo agency="ICPSR">04245</IDNo>
</titlStmt>
<serStmt>
<serName ID="Series3">American National Election Study (ANES)
Series</serName>
</serStmt>
</citation>
</stdyDscr>
<dataDscr>
<var ID="V403" name="V043131" files="F1" dcml="0"
intrvl="discrete" source="producer" wgt="not-wgt" temporal="N"
geog="N">
  <location StartPos="740" EndPos="740" width="1"
source="producer" />
  <labl source="producer">
    M1a. War in Afghanistan worth the cost
  </labl>
  <valrng source="producer">
    <range UNITS="REAL" min="1" max="5" source="producer" />
  </valrng>
  <invalrng source="producer">
    <range UNITS="REAL" min="9" source="producer" />
    <item UNITS="REAL" VALUE="8" source="producer" />
  </invalrng>
  <sumStat type="vald" source="producer" wgt="not-wgtd">
    1187
  </sumStat>
```


Questions?

- Mary Vardigan
vardigan@umich.edu